

Reviewer: Carl Walters

Title of Paper: PATH Final Report for Fiscal year 1998

Comments:

a) scientific soundness

I am beginning to think of the general PATH decision analysis approach as 1) developing a life cycle model of the form $N_t = N_{t-1} \exp(-Z_t)$ (suppressing details of age structure effects for ease of visualization), 2) decomposing total mortality rate Z_t into components that satisfy two “constraints”: consistency with historical N_{t+1} vs N_t data (stock recruitment parameters), and consistency with alternative hypotheses about impacts of past hydrosystem development and management and possible changes in “outside” (marine) mortality factors.

I continue to consider this a sound approach, subject to concerns mentioned in previous reviews about: (1) confounding of temporal effects on Z due to outside factors, with effects attributable to hydrosystem changes: in particular, attribution of negative Z anomalies after 1983 to passage effects in both CriSP and FLUSH, where in fact similar changes occurred for other chinook stocks outside the Columbia (eg Fraser, see Bradford 1994 CJFAS 51:965); and (2) whether the stock set used for analysis is in fact representative of the system or instead just “high-grades” the most productive (and currently most numerous) substocks.

I have become even more concerned about point (1) after seeing your comparisons of passage model predictions of reach survival to observed reach survivals. Model performance is not exactly stellar (does not capture the observed range of variation in survival), and this raises suspicion about whether the models correctly represent even the main historical drivers of passage survival: that is, there is a potential circularity, in that the models may have been “tuned” to historical data in the first place by seeking flow/temperature effects that would produce observed overall survival declines after the early 1980s. This means that not only is there an uncertainty along the CriSP/FLUSH assumptions dimension, but also about whether *either* model is even in the right ballpark in predictions of reservoir and whole system survival rate (I think this concern applies to both chinook types, and obviously also to any extrapolation from chinook to steelhead).

What this concern means is that I no longer trust your assessments about the range of uncertainty in recovery predictions under alternative policies, in particular I do not trust your finding that there is a very high long term recovery probability under the dam removal options (where the passage models become the dominant factor causing differences among policies in predicted performance). I suspect that after some reflection, you are going to have to admit considerably greater uncertainty about whether even these extreme measures will do the job. This concern is reinforced by the results in Table 1.3.6-1, which predicts considerably higher SAR than historical even for option A1. This amounts to concluding that actions since 1995, for which little empirical

evidence of effect is yet available, have already been quite effective at improving survival rates (or that there is a high probability of ocean survival being better in future).

b) general suitability of data

It appears that the PATH team has done an excellent job in bringing together auxiliary (e.g. life stage component survival, mortality rate) data, and in incorporating this information in the retrospective and prospective analyses. Except for the basic question raised above about whether the passage models are even in the right ballpark, I think the data generally support the range of hypotheses that you have admitted to the decision analysis.

You have very clearly and emphatically emphasized the gross “hole” in all life cycle models for the Columbia, namely the inability to clearly partition juvenile mortality within the system from mortality occurring below Bonneville but potentially impacted by what the fish have seen while in the system (e.g. transportation).

c) validity of conclusions/inferences

As noted above, I suspect the prospective analyses are too optimistic about 1) range of possible outcomes, 2) efficacy of recent (1995+) management, and 3) impact of dam removals. This optimism is at least clearly presented, in Fig. 1.3.6-3, which shows essentially no chance of continued decline.

d) suggestions for improvement/extension

The following may help make future analyses more credible and conservative:

- (1) you need not compute $P(\text{meeting goals})$ by doing many runs under every hypothesis combination. If you admit many more hypotheses (e.g. intermediate survival rates, D values) and choose the hypothesis at random for each run with probability proportional to your prior for that hypothesis, then a smaller number of runs/action can be used to estimate the overall means and probabilities of alternative goal outcomes. That is, you can sample simultaneously from hypotheses and random future anomalies when calculating policy performance. This simplification would prevent doing your very useful CART trees, but would make it much easier to evaluate/compare a lot of experimental policy options defined by treatment regimes and rules for responding to monitoring data gathered during such regimes (i.e. for evaluating experimental policy options).
- (2) You should not assume increasing abundance of some stocks following habitat enhancement would automatically trigger harvest rate increases; wouldn't continued application of ESA be used to prevent such increases?
- (3) Further on habitat effects, why assume such effects will be on Ricker a parameter rather than on habitat capacities (b parameter)? Effects may be on effective habitat size rather than density-independent components of f.w. survival.
- (4) In section 2.3.3, it appears that you have simulated additional mortality incorrectly. Shouldn't you look at effect of many successive years of 10%, 40%, etc. rather than

varying mortality randomly within the range over years (equivalent to just assuming around 25% every year).

- (5) Section 3.1.1: predator and habitat studies are not likely to provide credible “mechanistic” estimates of impact as claimed (who wrote this?). Why haven’t you compared survival rates predicted from the predation analyses with direct assessments from reach survival experiments? How do data in Fig. 3.1.1-1-2 compare to predictions from passage models, especially considering how much more variation is shown in Fig. 3.2.1-2 than predicted by the passage models? It looks like the predator “activity coefficient” alpha is just a poorly defined parameter (units not clear, but presumably mortality rate generated per mile of travel per unit predator density) that can be “tuned” by fitting the predator model to some independent source of survival data. If you are going to use the predator studies/data, at least make a serious effort to convert the data into genuine prior predictions of mortality impact based on proper dimensional analysis of the rate parameters.
- (6) The presentation on p 100-101 is confusing. You do not explain whether “recruitment” is inclusive of natural mortality occurring over ocean ages (i.e. total recruitment resulting from cohort analysis), or instead the “adult equivalent” estimate based on expanding river numbers only by the inverse product of age-specific ocean survival rates. Probably won’t make much difference, but you need to check this considering how much ocean harvest rates have changed in recent years.
- (7) As noted above, Fig. 3.1.2-10 shows survival rate index temporal patterns (declines after 1983) similar to non-Columbia chinook stocks. Your later residuals presentation (Fig. 3.3.2-3) indicates you attributed these patterns largely to passage model predictions, by assuming STEP changes started either 1970 or 76. Admitting a later possible start, and/or trend in effect rather than step effect, would force you to admit even worse confounding of passage and ocean effects in the R/S overall data, considerably greater range of uncertainty about efficacy of dam removal options.
- (8) On p 136, not clear why you took Deschutes residual series as a leading known, with residuals for others measured as departures from this. Why not just estimate a mean-zero shared time effect for all stocks, deviations from that? Would it make any difference (doubtful)?
- (9) In general, I find the approach to steelhead (section 4.2) unconvincing. Wouldn’t it be better to just admit a very wide range of uncertainty due to lack of historical data, supporting the need for experimental tests of policy options? Why get yourselves tangled in debates about the validity of your methods for making what amount to just guesses based on chinook data?

e) opportunities for integration of component analyses

As noted above, I think the main weakness in your prospective analyses at present has been to place unwarranted faith in component submodels (passage models) that in turn are derived from pretty sketchy real survival data and a whole raft of assumptions that cannot be tested from component data available. Your efforts at further integration should probably concentrate not on getting more bits of data and opinion, but rather on a broader appraisal of possible outcomes based on a more critical attitude toward all the model survival components.

f) relative priorities for future work

I agree that your top priority now should be the development and evaluation of alternative experimental designs. You have sketched out some design alternatives, and the trick mentioned above can be used to develop an efficient screening procedure for design alternatives. Your main emphasis in design analysis now should be on careful modeling of future data gathering and how to represent how that data will be analyzed and interpreted. In particular, you should pay considerable attention to whether any experimental response measures besides net escapement change should be used as indicators of response. All my intuition is that experimental results should be judged only in terms of net abundance response, with auxiliary data collection (and expense) justified only in so far as it may help split more detailed alternatives.

Addendum by email

Had a chance to look more closely at section 6 (experimental management), and have two additional (both nasty) comments to add to my review report:

1) Figure 6.2-2 (adaptive mgmt vs basic research) is deeply, fundamentally misleading. In complex dynamic systems, "basic research" does not "maximize learning"; such research cannot in principle deal with the kind of deep confounding of effects that you demonstrate by excellent example in Fig. 6.3.1. Research can help in hypothesis generation, and in posterior explanation of experimental response patterns, but it is simply wrong to claim that it can in any way substitute for actually seeing the responses to be explained! The correct distinction is just active/passive, and here the real issue is whether passive approach "hides" (or fails to test, or fails to reveal) responses, ie whether it prevents learning.

2) The data imply strong trends in total mortality rate Z_t that cannot confidently be attributed to (corrected for) known factors like passage survival changes. This means that a) for "irreversible" treatments (like dam removal) there is a high risk that post-pre estimates of treatment effect would be wrong, ie high risk that real response is due to some Z factor other than treatment, and there is no way to "control" for this risk; and b) for reversible treatments (transportation, hatchery, flow, etc.) confounding of treatment responses with other factors possibly causing Z change can only be avoided by interspersing (blocking) treatment and reference comparisons, ie by regularly operating the system under a reference treatment option in order to measure changes over time in Z due to factors other than the management treatment. This requirement for temporal reference comparisons greatly increases the time needed for effective experimentation, and seriously calls into question your representation of experimental options as incremental vs reverse staircase. Actually, you cannot conduct a

staircase experiment at all for whole-system treatments like dam removal, since such experiments are defined not by treatments within an experimental unit over time but rather by starting treatment on different experimental units at different times (i.e. you are wrong to represent years as experimental units in defining what you call a reverse staircase design).

to: INT:dmarmorek@msmail.essa.com
cc: int:lwb.env@worldnet.att.net

Clarifications to Above Review

-----Original Message-----

From: Dave Marmorek
Sent: Wednesday, April 21, 1999 11:59 AM
To: walters@fisheries.com
Subject: RE: Path 1998 review- CLARIFICATION {urgent}

Carl: Several people in PATH have requested a clarification of two of the paragraphs in your review. The questions follow your paragraphs.

PARAGRAPH 1

"I have become even more concerned about point (1) after seeing your comparisons of passage model predictions of reach survival to observed reach survivals. Model performance is not exactly stellar (does not capture the observed range of variation in survival), and this raises suspicion about whether the models correctly represent even the main historical drivers of passage survival: that is, there is a potential circularity, in that the models may have been "tuned" to historical data in the first place by seeking flow/temperature effects that would produce observed overall survival declines after the early 1980s. This means that not only is there an uncertainty along the CriSP/FLUSH assumptions dimension, but also about whether *either* model is even in the right ballpark in predictions of reservoir and whole system survival rate (I think this concern applies to both chinook types, and obviously also to any extrapolation from chinook to steelhead)."

CLARIFICATION QUESTION ON PARAGRAPH 1

Regarding "*comparisons of passage model predictions of reach survival to observed reach survivals*", were you referring to the fall chinook graphs on page 122 of the FY98 report (poor fits), or the spring summer chinook comparisons in the Weight of Evidence document (pg. 41-45; better fits) or both?

Carl Walters' reply:

I was referring to falls in particular, and most especially to the apparent ability of the passage models to "explain" post 1983 survival (R/S) trends that I suspect are shared at quite large geographic scales. I recognize that the poor fits of passage models to measured reach survivals (pg. 122) are partly due to comparing models to late releases, but the models claim to deal with flow/temperature effects of such timing.

PARAGRAPH 2

"What this concern means is that I no longer trust your assessments about the range of uncertainty in recovery predictions under alternative policies, in particular I do not trust your finding that there is a very high long term recovery probability under the dam removal options (where the passage models become the dominant factor causing differences among policies in predicted performance). "

CLARIFICATION QUESTION ON PARAGRAPH 2

What is your relative amount of trust in recovery predictions under transportation options and drawdown options?

Carl Walters' reply:

This one is easy: I don't trust any of the recovery predictions at all, under either option. This arises from seeing that none of the mean trajectory predictions would be for continuation of historical decline, meaning there is a basic (and quite possibly wrong) optimism somewhere in the survival calculation chain independent of passage models.

[Editor's note: This document provided 4/25/99 by Dave Marmorek for posting on the PATH web site.]